DOI:
10.1007/s11098-014-0406-9

Document Version
Peer reviewed version

EXPRESSING FIRST-PERSON AUTHORITY

(Forthcoming in *Philosophical Studies*)

Matthew Parrott
King's College London

(Penultimate Draft)

Abstract: Ordinarily when someone tells us something about her beliefs, desires or intentions, we presume she is right. According to standard views, this deferential trust is justified on the basis of certain epistemic properties of her assertion. In this paper, I offer a non-epistemic account of deference. I first motivate the account by noting two asymmetries between the kind of deference we show psychological self-ascriptions and the kind we grant to epistemic experts more generally. I then propose a novel agency-based account of deference. Drawing on recent work on self-knowledge, I argue that a person normally has a distinctive type of cognitive agency; specifically she is able to constitute her psychological attitudes by making judgments about what they ought to be. I then argue that a speaker expresses this agentive authority when she self-ascribes a psychological attitude and this is what justifies deferentially trusting what she says. Because the notion of expression plays a central role in this account, I contrast it with recent neo-expressivist theories.

When a person tells us that she believes that the Yankees are going to with the World Series or that she wants a beet salad for lunch, we ordinarily take what she says to be true. We may doubt whether the Yankees will win or find the taste of beets repulsive, but we immediately trust what an ordinary person says about her own psychological attitudes. In this way, what a person says about what she believes, desires, or intends seems to give us conclusive evidence about what she does believe, desire or intend, even in cases where the person speaking is a complete stranger. Indeed, we treat what a stranger says about her psychological attitudes in the same way we treat what our family or close friends say. Although familiar considerations ranging from pathologies to lies show that what someone says about her attitudes is never guaranteed to be true, when a person is sincere and engaged with her psychological attitudes in a normal first-personal

way, her assertions about them seem to enjoy a unique standing in conversations, a standing that appears to justify immediately deferring to her self-ascriptions.[1]

Deferring to psychological self-ascriptions is a way of conferring a special *epistemic* status on them. When we defer, we treat what a person says about her psychological attitudes as more likely to be true than what someone else says about them. As Austin (1946) noted, this can easily give the impression that we are responding to epistemic features of a speaker's assertions; that they somehow provide us with the best possible evidence as to the existence and character of her actual attitudes.[2] Why else would we be justified in deferring to what other people tell us if not because they are especially well situated to know about their own attitudes? If psychological self-ascriptions were not connected to some sort of epistemic privilege or advantage, why would they be so easily trusted?

---

[1] Philosophers sometimes use the term "avowal" to refer to an assertion made about one's own psychological attitudes. I am avoiding this term because it seems to me that different philosophers have different things in mind by "avowal". For example, Richard Moran uses it to refer to "a statement of one's belief which obeys the Transparency Condition;" (2001, pg. 101), which is a fairly narrow usage. Whereas Dorit Bar-On and Douglas Long think of "avowals" more broadly as any "present tense self-ascriptions of occurrent mental states." (2001, pg. 311) It is notable that neither of these is how Ryle understood "avowal" (his use of the term included self-ascriptions of moods, which he explicitly claimed were "not occurrences") (1949, pg. 83).

[2]Some may wish to contend that psychological self-ascriptions are not assertions because a speaker does not utter them with the intention of reporting her attitudes (cf. Bar-On 2004). This assumes that assertions are uttered only in order to report information, which is a restrictive conception of assertion (for more on the nature of assertion see MacFarlane (2009)). It seems to me that Williams is right when he says that "if a speaker comes out with a declarative sentence not as part of a larger sentence (as one might say, by itself) and there are no special circumstances, then he is taken to have asserted what is meant by that sentence." (2002, pg. 74) So, following Williams, I shall assume that openness in speech results in assertion regardless of whether the speaker's prior intention is to report some information.

Epistemic privileges explain why we deferentially trust medical doctors, mechanics, and other kinds of experts. We believe what they say about a particular set of facts because they are in the best position to know or discern the truth. They are, in this way, epistemic authorities with respect to a specific domain. It would therefore make a great deal of sense if the reason we deferentially trust what someone says about her psychological attitudes is that she is an epistemic authority with respect to them. Indeed, according to standard philosophical accounts, the reason we justifiably defer to a person's psychological self-ascriptions is because she makes them from some kind of epistemically authoritative perspective.

However, although deferring is a way that we confer a privileged epistemic status on a class of assertions, it need not be a response to any epistemic properties of those assertions, to those properties they have in virtue of their epistemic bases. In contrast to the standard approach, it may be that non-epistemic properties of psychological self-ascriptions justify deference. In the first section of this essay, I aim to motivate a non-epistemic view by highlighting two ways in which deference to psychological self-ascriptions looks to be different from the kind we confer on epistemic experts. As I shall argue, the presence of these two asymmetries suggests that deferring to psychological self-ascriptions may not be grounded in a speaker's epistemic authority or, for that matter, in any other epistemic properties of her assertions.

In section two, I shall argue that the first-person is fundamentally a standpoint of cognitive agency. From this point of view, an individual has a unique capacity to constitute her attitudes by judging what they ought to be. This, I claim, secures her a special kind of *agentive authority* for those attitudes. Although other philosophers have stressed the significance of agency to the first-person, they have focused heavily on its epistemic implications for securing self-knowledge, thereby overlooking the possibility that it might directly explain the phenomenon of deference. In section three, I shall

present a purely agency-based account of why we are justified in deferentially trusting psychological self-ascriptions. I shall argue that normally when a person self-ascribes a psychological attitude, she expresses her agentive authority to determine that attitude and thereby presents herself as a rational cognitive agent to others. Deferential trust is justified in this context because, in addition to expressing agentive authority, an individual's self-ascription semantically represents an attitude that she alone has the authority to constitute.

Since the notion of expression plays a central role in this agency-based account, it will be helpful to distinguish it from other forms of expressivism. Contemporary "neo-expressivists" hold that self-ascriptions directly express a person's underlying mental states and that this is why we are justified in deferring to them. In section four, I shall contrast the agency-based account presented in this essay with this more familiar form of expressivism. Neo-expressivists are right to notice that the deference we grant to psychological self-ascriptions is not based on their epistemic features but, as I shall argue, they fail to explain why this practice is justified.

Before continuing, I should also note at the outset that the account presented in this essay does not apply to sensations or other phenomenal states. These are in no way constituted or determined by a subject's cognitive agency; so any agency-based proposal will need to be supplemented in some manner in order to explain why we deferentially trust phenomenal self-ascriptions. However, there is some evidence that humans have discrete cognitive systems for attributing mental states to others, one responsible for agentive attitudes like belief, intention, and desire and the other for phenomenal states (cf. Gray et. al, 2011; Knobe and Prinz 2008). Thus, there is some empirical support for thinking that our justification for trusting what someone says about her own psychological attitudes may be different from the justification we have for trusting what she says about her phenomenal states. Naturally, the empirical data doesn't entail that

this is true but, since I lack sufficient space to argue for the claim in this essay, I shall simply assume that our reasons for deferring to phenomenal self-ascriptions are different in order to focus on ascriptions of psychological attitudes.[3]

1. Deferring to Authority

In a wide variety of contexts, it is appropriate to defer to what a person sincerely says about some subject matter.[4] In what follows, it will be important to keep in mind that although deferring involves trusting what someone says it is not merely trusting. To trust a speaker is to expect or depend on what they say to be true (cf. Faulkner, 2007). For example, I trust my friends in the sense that I expect them to tell me the truth and I therefore frequently rely on the truth of what they say. Since multiple individuals can be reliable sources of information, it is not uncommon to trust more than one person when it comes to a particular subject matter. Unlike this simple trust, deferential trust, as I understand it, is a comparative notion. When we defer to what a particular person says, we trust what she says about a particular subject matter more than what other speakers would say, even though those other speakers are also trustworthy. Deferring or deferentially trusting is thus a matter of treating the assertions of a particular individual (or group) as *more* likely to be true than those of another, who one nevertheless continues to trust. This is possible because trusting is often a matter of degree.

---

[3] This is not uncommon strategy among contemporary work on the authority of self-knowledge, cf. Moran (2001); Bilgrami (2006); Boyle (2009).

[4] In any discussion of deference, we must assume the speaker is sincere. In any context, if someone lies or is otherwise misleading, she loses her entitlement to deference. This is not to say we might not continue to defer to what she says, only that we are wrong to do so. In the remainder of this essay, I shall assume speakers are sincere.

To take an example, I deferentially trust what my mechanic says about the engine of my poorly performing car. I immediately take what he says about the engine to be true. This doesn't imply that I don't also trust my neighbor, who also makes comments on my car's engine, only that I *defer* to my mechanic--I treat what he says about the car as more likely to be true than what my neighbor says or would say. Assuming that both are completely sincere, both are deserving of my trust, but, in virtue of his epistemic expertise, my mechanic is entitled to deferential trust; he is entitled to my trusting what he says to a greater degree.

We often defer to epistemic experts, to individuals who are, in fact, especially well placed to know about certain facts or to learn about them in some way. I justifiably defer to what my mechanic says about my car because he is an epistemic authority on the car's engine. He both knows more and is in a much better position than I am to learn about this particular domain of facts. In deferring to what he says, I am responding to this epistemic property of his assertion, to the fact that it is made by someone with relevant expertise, someone who stands in an especially good epistemic position.[5] Because the epistemic properties of his assertions are sufficient to justify my deferentially trusting my mechanic, let's call this epistemic deference.

---

[5] According to some views of testimony, epistemic properties would not be what justify trusting my mechanic. Tyler Burge thinks that we are entitled *a priori* to accept any intelligible speech unless there are good reasons not to (Burge (1993)). If Burge is right, it will be more challenging to distinguish the kind of deference we grant to psychological self-ascriptions form the kind we bestow on epistemic experts. It will not, however, be impossible. Even if the justification in both cases were *a priori*, there would still be an asymmetry in the way that justification could be defeated. As I shall argue, explicit appeals to evidential support sometimes defeat our justification for deferring to psychological self-ascriptions but they never defeat our justification for trusting epistemic experts. I think this may be sufficient to distinguish the two, on the assumption that Burge is right. However, I also agree with the criticisms of Burge's account of testimony leveled by Fricker (1994) and Faulkner (2007).

There are at least two ways in which the kind of deference we practice toward psychological self-ascriptions looks different from this purely epistemic kind. First, consider the way that an explicit appeal to supporting evidence functions in a standard case of epistemic deference. When an epistemic expert makes an assertion, she can, and sometimes does, support it by explicitly citing further evidence. For instance, I might appear to be in a skeptical mood when I go to pick up my car, noticing this, my mechanic might appeal to some additional evidence to back up his claim that the timing belt is broken. He might say, for example, that 90% of Subaru's need a new timing belt after 150,000 miles or that the other mechanics in the garage agree with him. The more supporting evidence he appeals to, the more reason I have for believing his initial claim. Generally, when an epistemic expert makes an appeal to supporting evidence, the evidence strengthens the credibility of his assertion.

Things seem different when it comes to self-ascriptions of psychological attitudes. Even though a speaker does not typically base these on behavioral evidence, there is nevertheless almost always some available. One may therefore easily appeal to supporting evidence for a psychological self-ascription by taking up a third-person point of view. However, when someone does this, we are usually less inclined to trust what she says. In fact, her explicit appeal to supporting evidence for claims about what she believes, desires, or intends seems to undermine her entitlement to deference.

Suppose I tell you that I believe someone in my department is out to get me and that I think this (about myself) because (i) my long-time therapist has diagnosed me as paranoid and (ii) while at work, I have noticed myself furtively watching colleagues. How should you respond to my self-ascription of this belief? Should you trust that I do, in fact, believe someone in the department is out to get me? Should you defer to what I say about my beliefs over what my therapist says about them? The evidence I appeal to seems good enough; indeed it is evidence that you might have become aware of

independently of anything I said. But, in this case, it seems more natural for one to be hesitant and not take what I say to be true. Perhaps you might ask me what reasons I have for thinking a colleague is out to get me but I might confess to having none. That is, I might admit that there is no evidence someone *is* actually out to get me, but plenty for thinking it is something I believe. In these cases, appeals to supporting evidence seem odd and deferentially trusting what someone says looks unreasonable. In fact, it looks like the more a person explicitly appeals to behavioral evidence, the less justification we have for deferentially trusting what she says.

From an epistemic perspective, this is very difficult to understand. Generally, when a person cites more evidence for an assertion, it either becomes more credible or, at worst, there is no effect on the justification for trusting what she says. So why would a speaker's citing behavioral evidence in support of her self-ascription ever undermine our justification for trusting what she says?

One might reply to this question by emphasizing cases in which citing supporting evidence does appear to strengthen a speaker's entitlement to deference, for example, a case in which someone looks to behavioral evidence to support a self-ascription of a feeling or mood, as when I say that I must be afraid of spiders because my heart races and my palms sweat whenever I see one. This reply, however, misses the point. The question is why a speaker's appeal to supporting evidence would *ever* undermine her entitlement to deferential trust. From a strictly epistemic standpoint, it is very hard to see why there would be any case in which the justification for deferring to someone's assertions would be eroded by her citing additional evidence for the truth of what she said.[6]

[6] Continually citing supporting evidence for a claim will no doubt strike us as bizarre and this might lead us to think there is something psychologically amiss with a speaker, but the additional evidence doesn't undermine the epistemic status of what is said.

The second way in which the practice of deferring to psychological self-ascriptions looks importantly different from the kind we show to epistemic authorities is that in the latter case it seems that the relevant authority can be easily shared. When it is, the corresponding entitlement to deference is also shared. Thus, there may be two mechanics working in the garage with the same level of expertise on Subaru engines. If so, I have no more justification for deferentially trusting one over the other when it comes to the status of my car's engine. Because each is an epistemic authority to the same degree, they are both entitled to the same degree of deference, even if I happen to actually trust one more than the other. By contrast, it looks like, no matter how much epistemic expertise someone else has on my attitudes, another person's assertions about them cannot be entitled to deference to the same degree as my own.

Something along these lines is what A. J. Ayer (1963) has in mind when he asks us to imagine someone who, through telepathy, acquires expertise on the existence and character of another person's attitudes. Ayer's thought is that if someone had telepathic access to my beliefs, desires, and intentions, her assertions about them would be as reliably accurate as my own. In that case, however, wouldn't a listener be justified in deferring to what *either of us* says about them? Ayer claims she would not because he thinks that, when it comes to my attitudes, I am "the final authority concerning their existence and character" (1963, pg. 68). But why would that be true? Why would my assertions be entitled to deferential trust over someone else's, someone who is in just as good of an epistemic position as I am? Once again, from a strictly epistemic point of view, it is unclear why the two of us would not both be "final" authorities on my attitudes. Perhaps Ayer's notion of telepathy is far-fetched, but we don't really need it to make the point. We can imagine a person acquiring a sufficient level of expertise on my attitudes in an entirely third-personal manner, perhaps an extremely dedicated social psychologist or therapist who spends a great deal of time with me. The fact remains that,

unlike the case of the two mechanics, even if another individual were to have the same degree of epistemic expertise on my psychological attitudes and we therefore were both epistemic authorities, we would not be equally entitled to deference.

It is important that we not overestimate the significance of these two asymmetries. The discussion so far is intended only to motivate a non-epistemic approach to understanding the phenomenon of deferring to psychological self-ascriptions. It is not a conclusive argument against the traditional epistemic approach. Indeed, there are ways in which a defender of the epistemic approach might try to accommodate these asymmetries. Most obviously, many philosophers have claimed that a person *necessarily* enjoys an epistemically superior perspective on her attitudes. Accordingly, one might think the reason appeals to supporting evidence seem inappropriate is that a subject necessarily stands in a better position to know about her own attitudes. Likewise, one might argue that the reason a subject cannot share her entitlement to deference with others is that no one else could possibly acquire an equivalent level of expertise.

Because my aim in this section is only to motivate a non-epistemic account of deference, I shall not quarrel with this line of thought. There is, however, a worry for any view that holds a subject necessarily has epistemic authority for her attitudes, a worry that also gives a bit more momentum to developing a non-epistemic approach. There is a significant amount of psychological research that seems to show people are quite often not that good at discerning their own attitudes and, more importantly for this discussion, that third parties, especially close friends and family, are frequently in as good a position to know about them.[7] If things are as this data suggest any actual epistemic authority a

[7] The kind of data I have in mind can be found in various places: see, for instance, Carruthers (2010); Haybron (2007); Valzire (2010); Wilson (2002); and Wilson and Dunn (2004). The spark for much of this work was Nisbett and Wilson's paper (1977).

person may have with respect to her attitudes is certainly not necessary. To the extent that research indicates we sometimes lack epistemic authority for our attitudes, it will be difficult to explain the practice of deferring to psychological self-ascriptions in epistemic terms.[8] By contrast, if the reason we defer to what an ordinary speaker says about her attitudes is not an epistemological one, then any research demonstrating our epistemic limitations would not threaten the practice.

2. Agentive Authority

How might we develop a non-epistemic explanation of why we deferentially trust what people say about their psychological attitudes? In this section, I shall argue that the first-person point of view is fundamentally a perspective of cognitive agency. This will then lead to a proposal in the following section that explains the practice of deference on the basis of that agency.

We can most clearly observe the sort of cognitive agency I have in mind during critical deliberation. For example, when deliberating about whether or not to believe *P*, a rational subject will consider reasons for and against the truth of *P*. If she judges that reasons conclusively favor *P,* she thereby judges that she ought to believe that *P*.[9] Crucially, from the first-person standpoint, there is no additional step that she must take in order to ensure she actually does believe that *P*. When she relates to her beliefs in a first-personal way, her judgments about what she should believe immediately constitute what she does believe. Similar points about deliberation hold for other judgment-sensitive attitudes. In general, a rational subject deliberates on reasons peculiar to a

[8] This is why some philosophers have expressed skepticism about the epistemic privilege of introspection, cf. Schwitzgebel (2008), (2012) and Bayne and Spener (2010).

[9] It is important that these reasons are conclusive. Making a judgment about non-conclusive reasons in favor of *P* does not settle what one ought to believe.

specific attitude type in order to determine whether to form or maintain one. Deliberation would have no effect if a person's judgments about which attitudes she ought to have did not immediately determine which ones she did have.

Not all attitudes are formed or maintained by the activity of deliberation. But can we still be agents for things that we do not deliberate or act upon? Sydney Shoemaker claims we have many attitudes that we in fact never deliberate upon. As an example, he says "I believe that I am wearing pants…but it is hard to think of circumstances, other than those of a dream, in which it could be a question for me whether I believe this." (2003, pg. 396; cf. Heal, 2004) Some attitudes, like Shoemaker's belief that he is wearing pants, will never be subjected to reflection or re-evaluation. But if that is the only way in which we function as cognitive agents, then we lack agency with respect to many of our attitudes. After all, we have never engaged, and most likely will never engage in any sort of cognitive activity that affects their existence or character.

It is a mistake to think agency is present only where actions are; a person can be an agent even though she is not actually doing something. But Shoemaker has something stronger in mind when he confesses difficulty thinking of circumstances "in which it could be a question for me whether I believe this." Even if agents need not actually be acting, they surely must have a capacity to act and Shoemaker is suggesting that there may be no context in which he even could deliberate on his belief that he is wearing pants. If so, then there may be no possibility of his ever engaging in deliberative action, which might suggest he lacks the capacity for cognitive agency.

Shoemaker's worry is instructive because it shows how a common conception of cognitive agency is too restrictive. On this conception, we are agents only because we deliberatively evaluate our attitudes.[10] This activity, as it is usually understood, requires

[10] Burge (1998) draws heavily on this conception of agency in developing his account of self-knowledge. For a further discussion and criticism of Burge, see Parrott (ms).

one to attend to a question, like whether or not to believe that *P*. One's cognitive agency is exhibited, on this picture, by her considering and settling on a determinate answer to that question. But although this gives us perhaps the most vivid picture of a cognitive agency, it is not the only way in which we are such agents.

Instead, we can think of cognitive agency as set of dispositions possessed by a rational subject (cf. Parrott, ms). One of these would be the disposition to engage in deliberation when faced with an appropriate question or open possibility. But it also seems that there are other important agentive dispositions. For example, suppose, with Shoemaker, that my belief that *P* is one I never deliberate on. That is, for me, the question of whether or not *P* never arises. By believing *P*, I am nevertheless rationally committed to the truth of *P*. This means that if I were to become aware of some compelling evidence that *P* is false, I would immediately stop believing it. In this way, my belief is appropriately sensitive to my sense of the appropriate evidential reasons. It is important to notice that this need not be the result of my *deliberating* on whether or not *P*. Like Shoemaker, I believe that I am wearing pants and cannot even imagine myself reflecting on whether or not I am. But I would nevertheless immediately stop believing it were I to recognize that I was not clothed. As a rational subject, I am disposed to respond to reasons for my belief and those beliefs will cease to exist when I judge that there are none. This is how the existence and character of those attitudes depends on my cognitive agency even when I do not actually deliberate. They depend on my ability to assess reasons for or against those attitudes.

Someone might object that this less restrictive picture is not really a kind of cognitive agency. Is an agent really responsible for exhibiting dispositional responsiveness to actual and possible reasons? A psychological attitude changing in response to a reason is not obviously in the realm of agency. Indeed, for many of our most basic attitudes, it seems that the reasons for them simply assail us, often to the

point that we cannot help but have them. The problem with this objection, however, is that it fails to fully appreciate what it means to say a rational subject's attitudes depend on *her* reasons for or against them. The subject is the one that must judge, often implicitly or tacitly, that something or other is a conclusive reason for a particular attitude--that something *is* such a reason is not obvious nor is it independent from everything else the subject believes (cf. Raz, 1997). This process need not be deliberative or self-conscious, but it remains true that unless the subject herself has a capacity to assess reasons for or against her attitudes, those attitudes will not be appropriately sensitive to reasons.

I believe this conception of cognitive agency is the best way to interpret philosophers who speak of psychological attitudes as "commitments". Richard Moran, for instance, claims that when we relate to our attitudes in a first-personal way, we cannot avoid a "connection between the question about some psychological matter of fact and a commitment to something that goes beyond the psychological facts." (2001, pg. 77) So, although my desire for a martini is a "psychological fact", from the first-person perspective it is also a "commitment" to the fact that drinking a martini is, in some way or other, good or valuable, which is how it "goes beyond the psychological facts". As Moran notes, it is for this reason that a person can reasonably be criticized for having a desire on the grounds that the object of that desire is in no way good or valuable. Similarly, although my beliefs are "psychological facts", they are also my take on what is true. By believing that *P*, I take a stand on whether or not *P* is true independently of whether I happen to believe it.

When I relate to them in a first-personal way, my psychological attitudes are "commitments to something that goes beyond the psychological facts" because they can be sustained or extinguished exclusively on the basis of what I take to be good reasons for them. This is true even if you have a much better grasp of what are in fact good

reasons. Your appreciation of reasons will not immediately determine my attitudes. Only I have the capacity to constitute my belief or desire by recognizing that I ought to believe or desire something. This means I am uniquely a cognitive agent with respect to those attitudes. I am the only one who is in a position to *do* something, namely constitute their existence or character directly on the basis of reasons. Because I am the only one, having this capacity endows me with a special kind of *agentive authority* for my attitudes.

It is helpful to notice that one can have *agentive authority* in other situations. Consider an umpire at a baseball game. The umpire is an authority for whether the pitch that has just been thrown is a strike. This is because her judgments, and only her judgments, constitute strikes.[11] Notice that the umpire has this authority even when she is not in a better position than others to know whether the pitch is a strike. The umpire cannot be ignorant but loyal fans or computers are typically in just as good of, if not a much better, epistemic position for knowing whether a pitch is a strike. Their knowledge in no way undermines the umpire's authority precisely because it is not grounded in her comparative epistemic standing.

[11] Legend has it that former Boston Red Sox great Ted Williams once said that the strike zone is "whatever that day's umpire says it is". That is to say, the umpire does not merely discern whether a pitch is within some pre-determined area; the umpire's judgments constitute strikes. In saying this, Williams is rightfully deferring to the umpire's authority, but not because the umpire is in a better position to know what a strike is. Notice too that the constitutive relationship is slightly different in the umpire's case. Her judgments about whether *a pitch* is a strike constitute strikes. In the case of psychological attitudes, a person's normative judgment about what attitude she ought to have constitutes the attitude. This is obviously different from so-called 'constitutive theories' that attempt to explain self-knowledge in terms of constitutive relations between a person's first-order attitudes and her higher-order beliefs about them (e.g., Bilgrami (2006); Heal (2001); and Shoemaker (1994), (2012)). A full discussion of these views is beyond the scope of this paper (but see Parrott, ms)

The kind of agency we have been focusing on in this section has been the subject of much important recent work in philosophy. Unfortunately, because attention has been paid primarily to how a special form of self-knowledge might rest on it, there has been no attempt to understand how it might be connected directly to the practice of deferring to psychological self-ascriptions. Presumably, it would be thought that if agency secures a distinctive mode of self-knowledge, deference would be justified on the basis of this way of knowing. However, although clearly not in the traditional sense, this is essentially another epistemic explanation of deference. It shares with more traditional views the central idea that deferential trust is ultimately grounded in epistemic properties of a speaker's assertion, even though, in this case, the relevant epistemic properties are derived from a subject's agency. But if the previous section is right, the deferential trust we are trying to understand seems to be relatively insensitive to the epistemic properties of a speaker's assertion. So although I share the conviction that agency is a key to understanding the first-person, in my estimation its importance has not been stressed nearly enough.

3. Authoritative Expression

In order to competently engage in the self-conscious speech act of self-ascribing a psychological attitude, a speaker must have a sufficient grasp of the concepts involved in the content of her assertion. As we have seen, our concepts of the psychological attitudes are of states whose existence ought to be determined exclusively by good reasons. If that is right, then a person cannot competently ascribe an attitude to herself unless she has some understanding, at least tacitly, that it should be grounded in reasons. For example, if I truly report that I believe that *P*, I am thereby aware of having an attitude that I ought to have if and only if there are adequate reasons for the truth of *P*. This is

fundamental to what it is to believe that *P* and is therefore not something that a competent speaker can be wholly ignorant of.[12]

By self-ascribing such an attitude a speaker explicitly communicates her own take on what attitudes she actually has. Moreover, as we saw in the previous section, from the first-person perspective, a person does not merely recognize that the psychological attitude she actually has is one that she ought to have if only if adequate reasons support it, she also commits herself to there being such reasons (e.g., a rational subject who believes that *P* is thereby committed to there being adequate reasons supporting the truth of *P*).[13] If this is correct, then it is plausible that a person who is not alienated from her attitudes is at least tacitly aware of what her self-ascribed attitudes commit her to; for example, she is tacitly aware of being committed to the truth of *P* in virtue of explicitly asserting that she believes that *P*. If she were not aware of the commitments generated by her attitudes, then it would be wrong to hold her responsible in cases where she fails to live up to those commitments. It would seem wrong to criticize a subject for believing that *P* when evidence clearly indicates otherwise if she failed to understand that the attitude she self-ascribed committed her to *P* being true. Thus, if standing in a first-personal relation to an attitude involves being committed to certain things, we can think of self-ascribing an attitude from this perspective as communicating a speaker's endorsement of that commitment.[14]

---

[12] This awareness is not usually, nor need it be, explicitly represented.

[13] Thus we might think of the agentive character of the first-person in terms of it being the perspective from which one not only recognizes but also imposes obligations on one's self. A related discussion along these lines may be found in Chapter 12 of Soteriou (2013), in which he rightly notes: 'Your recognizing that you are under an obligation to do something is not sufficient for imposing that obligation on yourself, and so it is not sufficient for governing your conduct in the way that you think you should...' (pg. 305)

[14] It is worth emphasizing that these are only features of psychological attitudes that one relates to in a first-personal way. An attribution of a belief that *P* made on the basis of behavioral evidence does not

A speaker is only in a position to explicitly stand behind her psychological attitude and what it commits her to because she has agentive authority for the attitude. This is partly why an act of attitudinal self-ascription from the first-person perspective signals one's endorsement of what one's attitude commitments one to. The self-ascription is not itself the forming of any commitment but it communicates that the speaker's ratifies or endorses the commitment. She can do this only because she has the agentive authority to fulfill the commitments of her attitudes in virtue of having the capacity to constitute their existence and character in response to the right sorts of reasons. None of this means that a person must be explicitly aware or be able to explicitly represent the fact that her psychological attitudes involves these kinds of commitments, nor that she must be able to explicitly represent her agentive authority. But a rational subject must be tacitly aware of the fact that the attitude she is self-ascribing commits her to certain things, which she also has the authority to fulfill.

Assuming this is right, I would like to propose that an individual who has agentive authority with respect to her attitudes also expresses this authority when she self-ascribes them. This is a significant point and it helps to see that it is often true more generally. A person in a position of agentive authority with respect to a domain typically expresses that authority when making assertions about facts in that domain. Consider the umpire again. When she cries out "that is a strike", among other things, we can hear her express the authority to determine that the pitch is a strike. Similarly, when a military

---

commit one to the truth of *P*. Someone can quite easily attribute to herself the belief that her neighborhood is unsafe based on noticing how she walks nervously down the road while constantly looking over her shoulder without being committed to the truth of that belief (cf. Moran, 2001). Indeed, when one's self-ascription is made on the basis of third-personal data, one can even self-ascribe the belief that *P* while simultaneously judging that one ought to believe ~*P*. This is why Moore's paradox does not seem odd in cases of third-personal self-ascription.

general tells us about the location of his troops, it seems he expresses the authority to determine where they are based, or if the leader of a nation declares war on another country, his speech act will typically express his authority to make such a declaration. In a similar way, the proposal is that if a subject has agentive authority for her attitudes, she expresses that authority in making an attitudinal self-ascription.

Thus, we may say the following:

Authoritative Expression: For a speaker *a* and psychological attitude *M*, if *a* has agentive authority for *M*, *a's* assertion "I am *M*" expresses this agentive authority.

In self-ascribing *M,* a person is quite obviously not speaking about her authority; she does not represent it in the content of her assertion. Rather, we should envision this authority being expressed by what she says analogously to the ways in which a person might express her anger, resentment, joy, disappointment, reluctance, confusion, intelligence, or a sense of humor by saying certain things. A fairly wide range of properties can be expressed in speech without being explicitly represented. The expressive properties of a speech act are those that manifest a psychological property independently of its being based on a particular epistemic method or procedure. Typically, a speaker expresses a psychological property in virtue of the way in which she says something, in virtue of the manner of her speech act. For instance, a subject asserting *P* in one way will express her disappointment about *P* while her asserting *P* in a different way will express her nervousness about *P* (cf. Green, 2007, especially section 6.4). The specific aspect of a token speech act that functions to express the relevant psychological feature depends partially on contextual parameters. For example, sometimes anger is expressed by one's tone, other times by word choice. In order for agentive authority to be expressed, a speaker must relate to *M* in such a way such that her

take on the reasons for or against *M* can determine its existence and character (she must relate to it in a first-personal way), but otherwise we can reasonably expect the ways this authority is expressed to vary from one context to another.

By talking explicitly about my attitudes the content of what I say also semantically represents them to others. The truth conditions of any self-ascription may be represented as follows:

Self-Ascription: For a speaker *a* and psychological attitude *M, a's* assertion "I am *M*" is true iff (*a*)*M*.

The agency-based view I favor claims that the conjunction of *Authoritative-Expression* and *Self-Ascription* justifies deferential trust. If, as *Authoritative Expression* claims, my self-ascription expresses my agentive authority over my attitude, a competent listener will be in a position to recognize me as a cognitive agent with respect to that attitude, as someone who can meet the commitments of the attitude in the right way. Moreover, as *Self-Ascription* claims, she will understand that the truth conditions for what I say are the very same attitude that she understands me to have this authority over. Therefore, a listener is in a position to deferentially trust what I say because she understands that whether or not what I say is true depends on a fact that I have the unique authority to determine.

Even so, one might wonder whether speakers can really *express* this agentive authority. We are perhaps familiar enough with expressions of sadness, fear, or anger to allow that certain types of psychological states or conditions can be expressed, but we might nevertheless think there is a difference between these and properties like authority, one significant enough to make the latter inexpressible. However, I see no good reason to think agentive authority is something that could not be expressed. It is clearly different

from anger but that does not entail that it is inexpressible or that its expression would be especially difficult to recognize. Other structural properties, such as intelligence, stupidity, or melancholy seem to be expressible.[15] Or think, for example, of the different ways in which someone can express her confusion, or nervousness. Perhaps these cannot be voluntarily expressed, but that does not mean they are inexpressible (cf. Green, 2007). In any event, *Authoritative Expression* does not rely on the voluntary expression of agentive authority. Instead, it presumes a speaker expresses this authority involuntarily by voluntarily self-ascribing a psychological attitude which she understands commits her to certain things. Thus, the speaker must intentionally do something; she must consciously self-ascribe and thereby endorse the commitments of attitude *M*; so an expression of one's authority is not a completely passive phenomenon. But this does not mean that she voluntarily or willfully expresses her agentive authority.

A much more troublesome issue for this agency-based view is whether an appeal to some notion of expression is even necessary. For *Authoritative Expression* to be doing any work, saying that someone expresses her authority must not be equivalent to saying

[15] Someone might object that intelligence is exercised rather than expressed. I think this is wrong. If you are unconvinced, consider stupidity. Can someone exercise stupidity? It seems far more natural to me to say that stupidity is sometimes expressed by what a person says. Perhaps, one could say that in theses cases stupidity is exhibited but I take this to be equivalent to saying it is expressed (*Objection*: Someone exhibits stupidity because it is a feature of the *content* of what she says and so not something that is expressed. *Reply*: Imagine a very stupid person reciting the Peano axioms). This naturally raises the question of what the difference is between saying that a property is exhibited and saying that it is expressed. On my view, expressions are a subset of exhibitions, which are ways of making some internal condition manifest. However, some exhibitions are not expressions, partly because they do not involve a psychological or mental state (cf. Martin, 2010). For instance, I regularly exhibit my clumsiness but it seems odd to say that I express it. A complete account of the expressive would explain why our concept of the expressive is tied so closely to psychological or mental states but this is a very large topic beyond the scope of this essay. For some further discussion, see Green (2007) and Martin (2010).

that she has it.[16] But it might be thought that we are justified in deferring to what an umpire says simply because she just *is* the relevant authority. Why would the umpire also need to express that authority in some way? Similarly, we might think that a speaker only needs to *be* an authority in order for her self-ascriptions to be entitled to deference.[17] Might *Authoritative Expression* not be an unnecessary extravagance?

The answer to this question hinges entirely how we think about our ordinary practice of deferring to psychological self-ascriptions. Earlier we saw that whenever a person takes a third-person point of view on her attitudes, what she says about those attitudes does not seem to be entitled to deferential trust. If this is the right way to characterize the phenomenon, then even in cases where a speaker's third-personal self-ascriptions are *true*, they are not entitled to deferential trust. Notice that this does not mean they should not be trusted or believed at all, only that they should not be treated as more likely to be true than what some other person (like her therapist) might say about the speaker's attitudes.

But now consider two self-ascriptions of 'I am in *M*', one made from the first- and the other from the third-person perspective. Let's call them *a* and *b* respectively and stipulate that both are true. We might naturally think that they are the same kind of speech act, after all they have the same content and are both assertive. However, if they were the same kind of speech act, then they would both be equally entitled to deferential trust. So, if *b* is *not* entitled to deference whereas *a* is, it is because they are different kinds of speech act. But, since both self-ascriptions have the same content (i.e., 'I am in *M*'), the most plausible way to differentiate them is in terms of *a* also having additional

[16] It also must not be equivalent to saying she exercises it but we have already seen that exercising agency is not necessary for one to have agentive authority. This is because the attitudes for which we intuitively have authority include more than those we form through deliberation.

[17] cf. Burge (1996). I also think this is a plausible way of reading Moran (2001).

expressive properties that *b* lacks. That is to say, if we adopt the proposal that *a* also expresses some psychological condition of the speaker (i.e., her agentive authority) and *b* does not, we have a clear way of distinguishing the kinds of psychological self-ascription that are entitled to deferential trust from kinds that are not.[18] *Authoritative Expression* therefore offers a plausible explanation for how a listener could be in a position to reasonably differentiate *a* from *b* even in cases where both are true. It does so by suggesting that a first-personal self-ascription, in virtue of expressing a speaker's agentive authority, is different in kind from a third-personal one.[19]

This is not to say that the expressive properties of a particular speech act are transparent or that one cannot be mistaken in identifying a first-personal self-ascription. Discerning what a speaker is expressing in a particular speech act requires a listener to have developed an auditory recognitional capacity that is sensitive to that type of expression. Having a recognitional capacity of this sort is partly a matter of biological endowment and partly a matter of acculturation. This is analogous to the sort of visual recogntional capacity that an observer must have developed in order to see a natural kind (e.g., a tomato or an elm tree) on the basis of its distinctive appearance (cf. Millar, 2008; 2014). Novices will be unable to recognize that a speaker is expressing certain psychological properties because they will lack the capacity to do so. But even listeners who have developed the necessary capacities will be liable to error, to mistakenly taking a speaker to be expressing her contempt or resentment in cases where she is not. Indeed

[18] The fact that in each case the speaker stands in a different relation to the attitude she self-ascribes will not help because the relation an individual stands in to the referent of her self-ascription does not plausibly individuate the kind of speech act.

[19] Another way to motivate this line of thinking is by reflecting on Moore's paradox. We might think that whereas first-personal self-ascriptions of Moore-paradoxical propositions sounds bad, their third-personal counterparts sound fine. For further discussion, see Moran (2001) and Shoemaker (1995).

much of social awkwardness results from precisely this sort of thing. So we should expect a listener, even one who has fully developed the capacity to recognize expressions of agentive authority, to occasionally mistake a speech act that does not express that authority for a different kind that does.

Nevertheless, for those who are reluctant to accept *Authoritative Expression*, there is a closely related view available. One might claim that, in virtue of a person's cognitive agency, deferring to any psychological self-ascription is warranted *prima facie* (cf. Burge, 1996). So rather than requiring a listener to be able to distinguish different kinds of self-ascription, we could say that deferential trust is justified as long as nothing defeats it. Deferring to a third-personal self-ascription might be undermined or overridden by further considerations, such as the information that a person's ascription was based on behavioral evidence, but, according to this line of thinking, one always has some justification for deferentially trusting any psychological self-ascription.

It is important to see that this alternative view maintains, contrary to what I have claimed, that listeners are entitled, at least *prima facie*, to trust any claim a speaker makes about her attitudes, even those she relates to in a third-personal way. We must ask why this would be true. If we think that the kind of cognitive agency that underwrites deferential trust is exclusive to the first-person perspective, and if we think a subject abdicates that agency when she takes a third-person perspective on her attitudes, then what reason could we have for deferentially trusting all of a person's self-ascriptions? The line of thinking under consideration does not offer us any alternative explanation of deferential trust but agrees that it is grounded in the features of cognitive agency that are fundamental to the first-person perspective. So why would the entitlement to deference carry over to claims made from a third-person standpoint?

One idea is that deferentially trusting self-ascriptions is simply what it is to treat someone as a rational subject and is therefore a background condition on social

interaction with others. Yet, as I have already suggested, this seems to mischaracterize our actual practice. Instead of deferring to everything a person says about her attitudes, it rather seems that we confer deferential trust only in cases where we think the speaker stands in a first-personal relation to her attitudes.

Furthermore, on any view that claims we are *prima facie* entitled to defer to any self-ascription, it is very hard to see why the entitlement would be defeated in a case where a person makes an accurate self-ascription. If, as this view insists, deference is *prima facie* justified even when the subject stands in a third-personal relation to her attitudes, it isn't clear why making that relation more evident to a listener would undermine the listener's justification. But if a speaker's explicit appeal to behavioral evidence does undermine our justification for deferentially trusting what she says, even in a case where what she says is true, it suggests not that justification is defeated or overridden in cases of third-personal self-ascription, but that there is no justification in these cases. A listener may nevertheless be excused for sometimes deferring to more than she should but her deference would not be reasonable unless she had some way of discerning the utterances that are entitled to it.

4. Neo-Expressivism

The agency-based account I have presented relies on the notion that a person's authority can be expressed by self-ascriptive speech acts but the general idea that a speech act can express a psychological property is not new. Neo-expressivism claims that we justifiably defer to what a speaker says about her attitudes because her utterances express those attitudes. Like the agency-based account in the previous section, neo-expressivism does not ground deference on the epistemic basis of an assertion. It also rejects the standard epistemic approach's central tenet that the reason we trust what someone says about her attitudes is that she is in the best position to acquire knowledge of them. But, in contrast

to the agency-based view, neo-expressivism claims that the only thing being expressed in a speaker's assertion is the same attitude she is talking about.

Historically expressivists took the utterance of a self-ascription to be a kind of verbal outburst, like a moan, groan, or a whimper. Outbursts are thought to merely express underlying conditions. For example, when a person moans, she expresses pain without asserting a semantically evaluable content, without intentionally acting. We can understand that she is in pain not because she has said or done something, but because her moan has expressively revealed it to us. The traditional expressivist thought of psychological self-ascriptions in this way—like moans they simply revealed attitudes and lacked truth-conditions.

There are obvious shortcomings with the traditional view. Plenty of evidence indicates that, unlike moans and groans, the contents of self-ascriptions function like those of an assertion. For example, they are synonymous with attributions made by others. When I say, "I believe that New York is a great place to live," I ascribe the same thing you do by saying, "He believes that New York is a nice place to live." Traditional expressivism cannot make sense of this semantic continuity nor can it make sense of other features such as their ability to serve as premises in valid inferences or antecedents of conditionals.[20] Contemporary neo-expressivism departs from the traditional view by emphasizing that nothing about the notion of an utterance expressing an attitude prohibits it from asserting a semantically evaluable content. [21]

Nevertheless, the neo-expressivist shares with the traditional view a negative thesis about the type of speech act a speaker intentionally performs when self-ascribing

[20] Cf. Bar-On (2004), Finkelstein (2003) and Wright (1998).

[21] Although not every contemporary expressivist identifies as 'neo-expressivist', I use the term to include any theory (such as Finkelstein, 2003) that that thinks an expressive speech act can also have semantically evaluable content.

an attitude. Dorit Bar-On, who has done much to develop neo-expressivism, makes this point as follows:

"The point of the subject's use of words is not to offer a descriptive report of her state, or to provide evidence for its presence, to inform someone about it. The subject's act of self-ascription may have no other point than to vent her frustration, shout for joy, give voice to her fear, air her idea, articulate her thought, let out her anger, and so on." (2004, pg. 243)

This is supposed to close off the possibility of an epistemic gap between what a speaker says and the attitude she is talking about, a gap that would require the speaker to have some sort of epistemic access to her attitude. According to Bar-On, when someone performs an expressive act, she is not reporting some fact. For this reason, there is no epistemic method or procedure lying behind her self-ascription that might be mistaken or malfunctioning.

Because no epistemic access supports an expressive act, listeners cannot subject a speaker's self-ascription to epistemic criticism or questioning. There simply is nothing for them to question. When someone groans or moans, we do not ask her for reasons or justification. If a psychological self-ascription is an expressive act in the same sense, it too will be immune from these epistemic assessments. This is something Bar-On frequently stresses:

"To the extent that we regard a subject as simply giving voice to the condition she self-ascribes, rather than, say, providing an evidence- or recognition- based report on her own self-findings, it should indeed seem inappropriate to ask after the reasons she has for the different aspects of the self-ascription she produces when avowing. To do so would be to betray a misunderstanding of the character of her performance." (2004; pg. 263)

Although it is true that an expressive act is immune from epistemic criticism, having immunity is not equivalent to being entitled to deferential trust. When we defer we are treating a self-ascription as true and as something that ought to be believed. There is more to this than simply refraining from criticism. This is a way of conferring a kind of privileged epistemic status on the assertion, a status with clear implications for what ought to be believed.

Neo-expressivism must therefore do more than appeal to something like "expressive character" if it wants to account for deference. It must provide some account of why listeners are justified in taking what a speaker says to be true. A virtue of Bar-On's work is that she acknowledges this. In her opinion, the supplementary explanation is provided by the semantics of self-ascriptions, specifically by the same fact captured by *Self-Ascription*. Bar-On describes it as follows:

"Self-ascriptive verbal expressions *wear the conditions they are supposed to express on their linguistic sleeve*, as it were. A linguistic utterance (and its analogue in thought) directly reveals the *kind* of condition it is intended to express through its semantic content. It contains a component that semantically represents the relevant kind of condition." (Bar-On, 2004, pg. 315)

All expressive acts reveal some psychological condition; that is simply what it is to be an expressive act. But, in addition to expressing a belief, my utterance of "I believe that New York is a nice place to live", also "wears" this belief on its "linguistic sleeve". It semantically represents the very same attitude it expresses.

Two features of self-ascriptions thereby come together in Bar-On's account to explain why listeners are justified in deferring. First, as we have seen, *Self-Ascription* is generally true; so competent listeners understand that the truth condition of a

psychological self-ascription is whether or not a speaker has a particular attitude. But, whereas the agency-based view appeals to *Authoritative Expression* to explain why listeners presume the truth conditions hold, Bar-On claims that the truth conditions of the speaker's utterance are directly expressed.

It is important to see how there are two distinct senses of "expression" at work in this proposal.[22] First, a person's speech *act* is characterized by Bar-On as expressing her attitude. But, additionally, the product of this act as a kind of semantic expression of a psychological attitude. Neo-expressivism can explain deference only if these two senses of expression match up, only if the proposition that a speaker utters semantically express or represents the very same attitude that her speech act expresses in the non-semantic or "action" sense.[23] Bar-On's account therefore needs the following symmetry principle:

Matching: For any speaker *a*, *a*'s assertion "I am *M*" (1) expresses her underlying psychological attitude *M* and (2) is true iff (*a*)*M*.[24]

Matching is important because it is possible to recognize a token speech act to be an expressive act, to be an act that serves to "vent" or "give voice" to an underlying attitude, without also presuming that what is said is true. Someone can "give voice" to an underlying attitude ψ and thereby express ψ in the action sense without also semantically

[22] This is something that Bar-On is very clear on. Cf. Bar-On (2004), Chapter 8, (2009) and (2010a).

[23] Bar-On (2004) finds Sellars's distinction between action-expression, causal-expression, and semantic-expression helpful for articulating her position. However, I do not see how the causal sense of expression is relevant to her view. Moans and winces might be thought to be causal expressions of their underlying conditions but Bar-On seems to want to downplay even this. It therefore seems to me that the action sense and the semantic sense are the ones that matter to her neo-expressivism, which is perhaps why she drops discussion of the causal sense in more recent work (cf. 2009 and 2010a).

[24] The second conjunct reiterates *Self-Ascription.*

representing ψ in the content of what she says. Moreover, it seems clearly possible for someone to "give voice" to ψ by semantically representing some *other* psychological attitude χ. Bar-On implicitly relies on *Matching*, claiming that when we hear a person utter a self-ascription "we take it that she *is* in the relevant condition--the condition that is semantically referred to by the self-ascription, which is *the very condition that would render the self-ascription true*." (2004, p.p. 316-317) That may be true but certainly this is what neo-expressivism should explain (cf. Brueckner, 2010).

One reason *Matching* could be true is if a speaker *made* it true by deliberately using the content of her psychological self-ascription to express the same attitude the content refers to. Bar-On, however, explicitly rejects this idea, which is just as well since it would make her view depend on the speaker's epistemic access to the attitude she intends to express. Instead, Bar-On claims that although her view requires "a person do *something* intentionally", it does not require that she "intentionally express a mental state." (2010a, pg. 56) Rather, she thinks a psychological attitude gets expressed because a speaker "gives spontaneous expression to a present state of hers *by* performing some intentional act…that doesn't have expression as its intentional aim."[25] (2010a, pg. 56)

But, if a speaker is not responsible for *Matching*, what is?[26] Since attitudes can be and typically are expressed by speech acts that do not semantically represent them, and

[25] From these remarks, it seems clear that Bar-On is working with a fairly broad conception of intentional action. Some philosophers have been reluctant to accept that conception, arguing that the only way a psychological attitude could be expressed by an intentional speech act is if the speaker intends to express that attitude. This is at the center of Boyle's (2010) critique of Bar-On but it is more charitable to Bar-On if we assume a broader conception of intentional action.

[26] Bringing in a causal sense of expression might be thought to help here and one might be tempted to offer a causal version of *Matching*. If underlying psychological attitudes *caused* assertions that semantically represented those states, it would be a way of explaining why we justifiably defer to those assertions. There are several problems with this. First, it is false. Psychological attitudes cause different sorts of

since self-ascriptions also do not seem to need to express the conditions they semantically represent, why does *Matching* hold only when a speaker relates to her attitudes in a first-personal way? Neo-expressivists not only need *Matching* to be true in the good (first-personal) cases, they also need it to be false in the bad cases, in the third-personal cases where deference has seemed inappropriate, especially because when I relate to my attitudes in third-personal ways, those attitudes are still *expressible*. It is not uncommon for a repressed belief or desire to be expressed by what someone says or does. So in principle it seems that any belief or desire, even those that are repressed, could be expressed by a self-ascription that semantically represents it.

The neo-expressivist may wish to deny this. For instance, David Finkelstein argues that, although unconscious attitudes can be expressed in behavior, we are "unable to express them by self-ascribing them." (2003, pg. 119) But why is that true? Finkelstein claims it has something to do with consciousness and that might be plausible if we thought a speaker needed conscious awareness of her attitudes in order to "express them by self-ascribing them". Yet that would mean speakers were voluntarily responsible for *Matching*, which we have just seen reasons to resist. However, if the awareness associated with consciousness is not necessary for expressing one's attitudes by means of self-ascription, it is far from obvious why an unconscious attitude, although expressible, would be inexpressible by means of a self-ascription.

The closest Bar-On comes to offering an explanation of *Matching* is alluding to an expressive capacity: "when ascribing to myself a state with content *c* in the normal way, I suggest, I exercise an *expressive* capacity, the capacity to *use* content *c* (rather than some

---

speech acts, not just ones that semantically represent them. Second, even if it were true, one would want some sort of explanation for the causal relation. I think that explaining why there is matching between the semantic sense and causal sense is even more challenging than why there is matching between the semantic sense and the action sense.

other content *c*') to articulate, or give voice to my present state." (2010b, p. 9; cf. 2009, pg. 67) Perhaps there is a plausible developmental explanation of this capacity to match semantic content with expressed attitude. We might think that small children learn to express attitudes in a variety of ways and that, over time, they come to learn that a very good way to express a particular attitude is by using a content that semantically represents the attitude. In this way, self-ascriptions might be thought to take over a role once played by more natural expressions of the attitudes.[27] Although not implausible, this sort of explanation would make *Matching* contingent. If we did not learn to use self-ascriptions as a means to "vent" or "give voice" to the same attitudes they semantically represent, deference would not be justified. This conflicts with the intuition that deference is an appropriate response to any possible self-ascription made from the first-person point of view, an intuition Bar-On shares (2004, pg. 125; cf. Bar-On and Long, 2001).

Perhaps there are other ways to explain *Matching*. But, rather than exploring them, it is worth noting why these same questions do not arise for the agency-based view. First, the agency-based proposal explicitly claims that a speaker has agentive authority for an attitude if and only if she relates to it in a first-personal way. For this reason, there is no possibility of the speaker expressing that authority for attitudes she relates to in a third-personal way, whether via a self-ascription or any other type of speech act. The relevant *expressible* feature is only exemplified when a speaker takes up the first-person point of view. By contrast, according to neo-expressivism, neither the relevant expressible properties, viz., psychological attitudes, nor the speaker's capacity to self-ascribe those properties are exclusive to the first-person perspective. So there is the explanatory question of why the two coincide only in certain cases.

[27] Bar-On offers a picture of what such an explanation would look like in Chapter 8 of her book. It is not clear to me whether she endorses it. A similar developmental account is defended in Green (2007).

Secondly, for the agency-based view, a subject's act of self-ascribing a psychological attitude is not individuated by the particular underlying attitude she 'vents' or 'gives' voice to.[28] This does not mean that in certain contexts a subject could not express a particular attitude, such as her desire for cake, by making a psychological self-ascription that semantically represents that desire. However, whereas the neo-expressivist account relies on *Matching* between the expressed attitude and the semantic content of the subject's self-ascription to explain deference, the agency-based proposal does not. Thus in any context where a subject's speech act expresses some attitude that is not the one it semantically represents, the speaker could still be entitled to deferential trust in virtue her standing in the appropriate first-personal relation to the semantically represented attitude. Perhaps these sorts of cases are infrequent but they seem possible (e.g., asserting 'I believe that you are being unhealthy by drinking so frequently' might express my desire for you to take better care of yourself). But since *Matching* is violated in these cases the neo-expressivist must deny the appropriateness of deference (or else re-describe the case).

I am sympathetic to neo-expressivism and the agency-based view is not incompatible with the idea that psychological attitudes have characteristic expressions. Indeed, I find it extremely plausible that a person's attitudes are expressed by what she says and does and am quite attracted to the idea that this fact grounds our ways of knowing about the attitudes of others. But explaining how we ordinarily come to know about another person's attitudes by means of their characteristic expressions is not the

[28] In the previous section I suggested that there are two different kinds of self-ascription, those made from the first-person perspective which express agentive authority and those made from the third-person perspective that do not. Yet the alternative view on which there is only one kind of self-ascription would also not individuate them in terms of the attitudes they expressed. So neither approach requires *Matching* to be true.

same as explaining our practice of deferentially trusting psychological self-ascriptions. Furthermore, there are many ways in which a person's attitudes can be expressed, most of which do not involve self-ascription. So if listeners are primarily responding to the expression of an underlying attitude, it would be strange for them to pay special attention to self-ascriptions.

Deferring to psychological self-ascriptions is a peculiar response to what a person *says* about her own attitudes. And, although there are various ways in which a person can express her attitudes, self-ascribing them is the paradigmatic way in which she informs us of her own take on her attitudes. The main proposal of this essay has been that it makes sense for us to show special consideration for a person's take on her own attitudes because, when she relates to them in a first-personal way, she is in a position to determine what those attitudes are.[29]

Bibliography

Austin, J. L. 1946: Other Minds. *Aristotelian Society Supplementary Volume* 20: 122-197.

[29] Earlier versions or parts of this essay were presented at the University of Oxford Theoretical Work in Progress Group, the University of Puget Sound, UC Berkeley, and UCLA. I am grateful to everyone who offered questions and comments on those occasions, particularly to Gregory Antill. For extremely thoughtful comments, and for many stimulating discussions on the topic of this essay, I must express special thanks to Andreas Anagnostopoulos, Tony Bezsylko, John Campbell, Anil Gomes, Nick Jones, Markus Kohl, Berislav Marusic, Michael Martin, Josh Sheptow, Michael Sollberger, James Stazicker, Barry Stroud, Lee Walters, and Daniel Warren. Finally, I wish to thank an anonymous referee for taking the time to read this essay and for offering very interesting and constructive feedback.

Ayer, A. J. 1963: Privacy. In his *The Concept of a Person and Other Essays*. London: Macmillan.

Bar-On, Dorit and Long, Douglas. 2001: Avowals and First-Person Privilege. *Philosophy and Phenomenological Research* 52: 311-335.

Bar-On, Dorit. 2010a: Avowals: Expression, Security and Knowledge: Reply to Matthew Boyle, David Rosenthal, and Maura Tumulty. *Acta Anlytica* 25: 47-74.

Bar-On, Dorit. 2010b: Neo-Expressivism: Avowals 'Security and Privileged Self-Knowledge. In Anthony Hatzimoysis (ed.) *Self-Knowledge*. Oxford: Oxford University Press.

Bar-On, Dorit. 2004: *Speaking My Mind*. Oxford: Oxford University Press.

Bayne, Tim and Spener, Maja. 2010: Introspective Humility. *Philosophical Issues* 20 (1):1-22.

Bilgrami, Akeel. 2006: *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press.

Boyle, Matthew. 2009: Two Kinds of Self-Knowledge. *Philosophy and Phenomenological Research* 78 (1):133-164.

Boyle, Matthew. 2010: Bar-On on Self-Knowledge and Expression. *Acta Analytica* 25: 9-20.

Brueckner, Anthony. 2010: Neo-Expressivism. In Anthony Hatzimoysis ed. *Self-Knowledge*. Oxford: Oxford University Press.

Burge, Tyler. 1993: Content Preservation. *Philosophical Review* 102:457-488.

Burge, Tyler. 1996: Our Entitlement to Self-Knowledge. *Proceedings of the Aristotelian Society* 96: 91-116.

Burge, Tyler. 1998: Reason and the First-Person. In Crispin Wright, Barry Smith and Cynthia MacDonald, (eds.), *Knowing Our Own Minds*. Oxford: Clarendon.

Carruthers, Peter. 2010: Introspection: Divided and Partly Eliminated. *Philosophy and Phenomenological Research* 80: 76-111.

Faulkner, Paul. 2007: On Telling and Trusting. *Mind* 116: 875-902.

Finkelstein, David. 2003: *Expression and the Inner*. Cambridge: Harvard University Press.

Fricker, Elizabeth. 1994: Against Gullibility. In A. Chakrabarti & B. K. Matilal (eds.) *Knowing from Words*. Kluwer.

Gray, K., Knobe, J., Sheskin, M., Bloom. P. & Barrett, L. F. 2011: More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology* 101(6): 1207-1220.

Green, Mitchell. 2007: *Self-Expression*. Oxford: Oxford University Press.

Haybron, Daniel. 2007: Do We Know How Happy We Are? *Nous* 41: 391-428.

Heal, Jane. 2001: On First Person Authority. *Proceedings of the Aristotelian Society* 102: 1-19.

Heal, Jane. 2004: Moran's *Authority and Estrangement*. *Philosophy and Phenomenological Research* 69: 427-432.

Knobe, Joshua and Prinz, Jesse. 2008: Intuitions about Consciousness: Experimental Studies. *Phenomenology and Cognitive Science* 7: 67-83.

MacFarlane, John. 2010: What is Assertion? In Jessica Brown and Herman Cappelen (eds.) *Assertion*. Oxford: Oxford University Press.

Martin, M. G. F. 2010: Getting on Top of Oneself: Comments on *Self*-Expression. *Acta Analytica*. 25: 81-88.

Millar, Alan. 2014: Reasons for Belief, Perception, and Reflective Knowledge. *Aristotelian Society Supplementary Volume* 88 (1): 1-19.

Millar, Alan. 2008: Perceptual-Recognitional Abilities and Perceptual Knowledge. In Adrian Haddock & Fiona Macpherson (eds.) *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press. 330--47.

Moran, Richard. 2012: Self-Knowledge, 'Transparency' and the Forms of Activity. In Declan Smithies and Daniel Stoljar (eds.) *Introspection and Consciousness*. Oxford: Oxford University Press.

Moran, Richard. 2001: *Authority and Estrangement*. Princeton, NJ: Princeton University Press.

Nisbett, Richard and Wilson, Timothy. 1977: Telling More than we can Know: Verbal Reports on Mental Processes. *Psychological Review* 8: 231-259.

Parrott, Matthew. ms: Self-Blindness and Self-Awareness.

Raz, Joseph. 1997: The Active and the Passive. *Aristotelian Society Supplementary Volume* 71 (1): 211–228.

Ryle, Gilbert. 1949: *The Concept of Mind*. Chicago The University of Chicago Press.

Schwitzgebel, Eric. 2012: Introspection, What? In Declan Smithies and Daniel Stoljar (eds.) *Introspection and Consciousness*. Oxford: Oxford University Press.

Schwitzgebel, Eric. 2008: The Unreliability of Naive Introspection. *Philosophical Review* 117 (2): 245-273.

Shoemaker, Sydney. 2012: Self-Intimation and Second-Order Belief. In Declan Smithies and Daniel Stoljar (eds.) *Introspection and Consciousness*. Oxford: Oxford University Press.

Shoemaker, Sydney. 2003: Moran on Self-Knowledge. *European Journal of Philosophy* 3: 391-401.

Shoemaker, Sydney. 1995: Moore's Paradox and Self-Knowledge. Reprinted in his *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press, 1996.

Shoemaker, Sydney. 1994: Self-Knowledge and 'Inner Sense'. Reprinted in his *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press, 1996.

Soteriou, Matthew. 2013: *The Mind's Construction*. Oxford: Oxford University Press.

Valzire, Simine. 2010: Who Knows what about a Person? The Self-Other Knowledge Asymmetry (SOKA) Model. *Journal of Personality and Social Psychology* 98: 281-300.

Williams, Bernard. 2002: *Truth and Truthfulness*: Cambridge: Harvard University Press.

Wilson, Timothy. 2002: *Strangers to Ourselves*. Cambridge: Harvard University Press.

Wilson, Timothy, and Dunn, Elizabeth. 2004: Self-Knowledge: Its Limits, Value, and Potential for Improvement. *Annual Review of Psychology* 55: 493-518.

Wright, Crispin. 1998: Self-Knowledge: The Wittgensteinian Legacy. In Crispin Wright, Barry Smith and Cynthia MacDonald, eds., *Knowing Our Own Minds*. Oxford: Clarendon.